We claim:

1. A server-side caching method for use with a server cluster including at least one central storage device and a plurality of servers having respective cache storage devices, the method comprising the steps of:

receiving a client request for a data object from a client device with one of the servers;

determining whether the data object is being cached by the server that received the client request;

determining whether a server that did not receive the client request is caching the data object in response to a determination that the server that received the client request is not caching the data object;

obtaining a copy of the data object from the cache storage device of a server that did not receive the client request in response to a determination that the server that received the client request is not caching the data object and a determination that the data object is cached in a server that did not receive the client request; and

transmitting the data object to the client device.

2. A method as claimed in claim 1, wherein the step of transmitting the data object to the client device comprises:

transmitting the data object from the server that received the client request to the client device in response to a determination that the data object is being cached by the server that received the client request.

3. A method as claimed in claim 1, wherein the step of transmitting the data object to the client device comprises:

sending the copy of the data object from the cache storage device of a server that did not receive the client request to the server that received the client request; and

sending the copy of the data object from the server that received the client request to the client device.

4. A method as claimed in claim 3, further comprising the step of:

sending at least one of data object request rate information and a data object replication recommendation to the server that received the client request with the copy of the data object.

5. A method as claimed in claim 1, wherein the step of transmitting the data object to the client device comprises:

sending a copy of the data object from the central storage device to the server that received the client request in response to a determination that the server that received the client request is not caching the data object and a determination that the data object is not cached in a server that did not receive the client request.

6. A method as claimed in claim 5, further comprising the step of:

caching the copy of the data object in the cache storage device of the server that received the client request.

7. A method as claimed in claim 1, further comprising the step of:

maintaining a cache status table including a list of data objects stored in the respective cache storage devices of the plurality of servers.

8. A method as claimed in claim 7, wherein the step of determining whether a server that did not receive the client request is caching the data object comprises:

querying the cache status table.

9. A method as claimed in claim 7, wherein the step of transmitting the data object to the client device comprises:

sending a copy of the data object from the central storage device to the server that received the client request in response to a determination that the server that received the client request is not caching the data object and a determination that the data object is not listed in cache status table.

10. A method as claimed in claim 9, further comprising the steps of:

caching the data object in the cache storage device of the server that received the client request; and

updating the cache status table to reflect that a copy of the data object has been cached in the server that received the client request.

11. A method as claimed in claim 1, wherein the step of obtaining a copy of the data object from the cache storage device of a server that did not receive the client request comprises:

obtaining a copy of the data object from the cache storage device of a server is operating below a predetermined load threshold.

12. A server system for use with a plurality of client devices, comprising;

a local area network adapted to be connected to a wide area network;

a plurality of servers connected to the local area network and having respective cache storage devices adapted to cache data objects; and

a cache load server process running on at least one of the servers that maintains a list of data objects cached in the cache storage devices.

13. A server system as claimed in claim 12, wherein each of the servers runs a load daemon process that monitors server loading and transmits server loading information to the cache load server process.

14. A server system as claimed in claim 12, wherein each of the servers runs a web server process that receives requests for data objects from client devices and queries the cache load server process to determine whether another server is caching a requested data object in response to a determination that the associated server is not caching the requested data object.

15. A server system as claimed in claim 14, wherein the web server processes queries the cache load server process to determine whether a

server that is caching a requested data object is below a predetermined server loading threshold.

16. A server system as claimed in claim 12, further comprising:

a dispatcher adapted to connect the local area network to the wide area network.

17. A server system as claimed in claim 12, further comprising:

at least one central storage device connected to the local area network.

18. A server system for use with a plurality of client devices, comprising;

a local area network adapted to be connected to a wide area network;

a plurality of servers connected to the local area network and having respective cache storage devices adapted to cache data objects; and

means for cooperatively caching data objects within the cache storage devices such that cached data objects may be shared by the plurality of servers.

19. A server system as claimed in claim 18, wherein each of the servers includes means for receiving a client request for a data object from a client device and means for determining whether the data object is being cached locally.

20. A server system as claimed in claim 18, wherein the means for cooperatively caching data objects includes means for determining whether a server that did not receive the client request is caching the data object.

21. A server system as claimed in claim 20, wherein the means for cooperatively caching data objects includes means for determining whether a server that did not receive the client request and is caching the data object is operating below a predetermined loading threshold.

22. A server system as claimed in claim 18, further comprising:

a dispatcher adapted to connect the local area network to the wide area network.

23. A server system as claimed in claim 18, further comprising:

at least one central storage device connected to the local area network.